

日 本 国 特 許 庁

PATENT OFFICE  
JAPANESE GOVERNMENT



別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日

Date of Application:

1998年12月16日

出 願 番 号

Application Number:

平成10年特許願第357023号

出 願 人

Applicant(s):

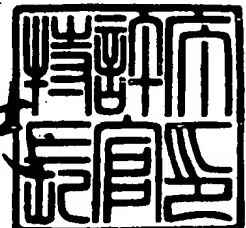
富士通株式会社

CERTIFIED COPY OF  
PRIORITY DOCUMENT

1999年 4月 2日

特 許 庁 長 官  
Commissioner,  
Patent Office

山 建 志



【書類名】 特許願

【整理番号】 9803605

【提出日】 平成10年12月16日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 17/60

【発明の名称】 表画像処理装置及びそのプログラム記憶媒体

【請求項の数】 8

【発明者】

    【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

    【氏名】 小原 敦子

【発明者】

    【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

    【氏名】 直井 聡

【特許出願人】

    【識別番号】 000005223

    【氏名又は名称】 富士通株式会社

【代理人】

    【識別番号】 100087848

    【弁理士】

    【氏名又は名称】 小笠原 吉義

    【電話番号】 03-3807-1151

【選任した代理人】

    【識別番号】 100074848

    【弁理士】

    【氏名又は名称】 森田 寛

【選任した代理人】

    【識別番号】 100087147

【弁理士】

【氏名又は名称】 長谷川 文廣

【手数料の表示】

【予納台帳番号】 012586

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9707817

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 表画像処理装置及びそのプログラム記憶媒体

【特許請求の範囲】

【請求項 1】 入力された画像から縦及び横方向の直線を抽出する直線抽出手段と、

前記直線抽出手段の抽出した直線の端点を起点とする斜め成分を抽出し、抽出した斜め成分を用いて丸角候補領域を検出する丸角候補領域検出手段と、

前記丸角候補領域検出手段の検出した丸角候補領域を含むセルを抽出するセル抽出手段と、

前記セル抽出手段の抽出したセルの角部について丸角部を決定する処理を行う丸角部決定手段とを備える

ことを特徴とする表画像処理装置。

【請求項 2】 前記丸角候補領域検出手段が、前記直線抽出手段の抽出した直線の内、縦方向の直線の端点を起点とする第 1 の斜め成分と、横方向の直線の端点を起点とする第 2 の斜め成分とを別個に抽出することにより、前記斜め成分を抽出する

ことを特徴とする請求項 1 に記載の表画像処理装置。

【請求項 3】 前記丸角候補領域検出手段が、

当該第 1 及び第 2 の斜め成分が互いに重なる場合に、当該部分が丸角候補領域であるとし、

当該第 1 及び第 2 の斜め成分が互いに予め定められた距離以下の位置関係にあり、かつ、両者の間に直線の特徴を示すパターンがあるか両者が互いに接触又は重なっている場合に、当該部分が丸角候補領域であるとし、及び／又は、

当該斜め成分の近傍に他の斜め成分が存在せず、かつ、当該斜め成分の先端に直線の特徴を示すパターンがある場合に、当該部分が丸角候補領域であるとし、前記丸角候補領域を検出する

ことを特徴とする請求項 2 に記載の表画像処理装置。

【請求項 4】 前記丸角部決定手段が、

前記セル抽出手段の抽出した当該セルの角部において画素密度が予め定められ

た順に変化する場合に、当該部分が丸角部であるとし、

前記画素密度の変化に基づいて決定された丸角部が存在する場合に、前記入力された画像の他の角部も丸角部であるとし、及び／又は、

当該セルを構成する前記直線抽出手段の抽出した直線の端点と端点とを結ぶように生成された  $n$  次関数からなるパターンと、前記入力された画像とがマッチングする場合に、当該部分が丸角部であるとするにより、前記丸角部を決定する

ことを特徴とする請求項 1 に記載の表画像処理装置。

【請求項 5】 当該表画像処理装置が、更に、

罫線により構成された帳票の画像を入力するための画像入力手段と、

前記丸角部決定手段により決定された丸角部を含むセルについて、当該丸角部を文字領域から除くことにより、当該文字認識領域を決定する文字認識領域決定手段とを備える

ことを特徴とする請求項 1 に記載の表画像処理装置。

【請求項 6】 入力された画像から抽出された縦及び横方向の直線の内、縦方向の直線の端点を起点とする第 1 の斜め成分と、横方向の直線の端点を起点とする第 2 の斜め成分とを別個に抽出することにより、前記直線を起点とする斜め成分を抽出し、

当該第 1 及び第 2 の斜め成分が互いに重なる場合、当該第 1 及び第 2 の斜め成分が互いに予め定められた距離以下の位置関係にありかつ両者の間に直線の特徴を示すパターンがあるか両者が互いに接触又は重なっている場合、及び／又は、当該斜め成分の近傍に他の斜め成分が存在せずかつ当該斜め成分の先端に直線の特徴を示すパターンがある場合、当該部分が丸角候補領域であるとするにより、前記丸角候補領域を検出する

ことを特徴とする表画像処理装置。

【請求項 7】 入力された画像から抽出された縦及び横方向の直線に基づいて抽出されたセルであって、丸角候補領域を含むセルの角部について、

当該セルの角部において画素密度が予め定められた順に変化する場合に、当該部分が丸角部であるとし、

前記画素密度の変化に基づいて決定された丸角部が存在する場合に、前記入力された画像の他の角部も丸角部であるとし、及び／又は、

当該セルを構成する前記直線の端点と端点とを結ぶように生成された  $n$  次関数からなるパターンと、前記入力された画像とがマッチングする場合に、当該部分が丸角部であるとする事により、

前記丸角部を決定する処理を行う

ことを特徴とする表面像処理装置。

【請求項 8】 入力された画像から縦及び横方向の直線を抽出する直線抽出処理と、

前記直線抽出手段の検出した直線を起点とする斜め成分を抽出し、抽出した斜め成分を用いて所定の処理を行うことにより丸角候補領域を検出する丸角候補領域検出処理と、

前記丸角候補領域検出手段の検出した丸角候補領域を含むセルを抽出するセル抽出処理と、

前記セル抽出手段の抽出したセルの角部について所定の処理を行うことにより丸角部を決定する処理を行う丸角部決定処理とを、

当該表面像処理装置であるコンピュータに実行させるプログラムを格納することを特徴とするプログラム記憶媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、表面像処理装置及びそのプログラム記憶媒体に関し、特に、丸角を有する表面像を正確に処理することができる表面像処理装置及びそのプログラム記憶媒体に関する。

【0002】

【従来の技術】

コンピュータの入力装置として文字認識装置（OCR）が多く用いられている。文字認識装置によれば、帳票等の予め文字を書く位置を指定した文書に対して、指定した範囲内に文字がきれいに書かれている場合、かなり高い認識率で文字

を自動認識することができる。なお、ここに言う文書は、枠等の罫線の色及び濃度が、ドロップアウトカラーではなく、黒等の文字と同一の色及び濃度であるものである。

## 【0003】

しかし、書かれた文字が少しでも指定した範囲を越えると、著しく認識率が低下する。例えば、文字が指定範囲を示す枠や罫線に接触したり、又は当該範囲からはみ出した場合である。

## 【0004】

そこで、位置や形式等が未知の帳票であっても、表形式枠中の枠を正確に抽出でき、文字と枠とが接触したり文字のはみ出しがある場合でも文字領域を正確に抽出できる技術が開発されている。

## 【0005】

## 【発明が解決しようとする課題】

様々な種類がある帳票の内、直交する2本の直線からなる角部ではなく、円弧のような丸い曲線からなる角部（以下、丸角又は丸角部）を有する表構造の帳票が存在する。丸角表構造の一例を図18に示す。この形式の帳票は現在使用されている一般的な帳票の中で大きな割合を占める。このため、文字認識装置等の表画像処理装置においては、丸角を認識できないと、帳票の処理に支障を生じる。従って、丸角の表構造を有する帳票の丸角の正確な処理のために、種々の提案がなされている。

## 【0006】

例えば、縦の罫線及び横の罫線（縦横罫線）を抽出した後、縦横罫線が相互に交差せずにある一定以下の間隔に位置している場合に、当該部分を丸角部であると認識する技術が提案されている（例えば、特開平7-282191号公報参照）。しかし、この技術によっては、第1に、画像が掠れた場合に丸角部を正確に処理できない可能性が高い。特に、当該丸角の部分が掠れてしまった場合には正確に処理できない。第2に、この技術を適用する前提として縦方向及び横方向の双方共に罫線を抽出できていなければならない。従って、この技術は、図18（A）に示す形式の帳票には適用できるが、図18（B）に示す縦方向の罫線が始

めから存在しない形式の帳票には適用できない。

【0007】

また、縦横罫線を抽出した後、縦横罫線が相互に交差せずにある一定以下の間隔に位置している場合に、当該部分と予め用意したパターンとのマッチングを採ることにより、角部（丸角部）の形状を判断する技術が提案されている（例えば、特開平7-14000号公報参照）。しかし、この技術によっては、第1に、予め多数のマッチングパターンを持つ必要があり、メモリ容量が極めて大きくなる。第2に、この技術も前提として縦方向及び横方向の双方共に罫線を抽出できなければならないので、図18（A）に示す帳票には適用できるが、図18（B）に示す帳票には適用できない。

【0008】

更に、縦横罫線を抽出した後、四方を罫線により囲まれた領域（セル）を抽出し、当該セルの内側の輪郭探索を行い、探索方向の変化によって角部が丸角部であるか否かを判断する技術が提案されている（例えば、特開平8-212292号公報参照）。しかし、この技術によっては、第1に、罫線に文字が接触していた場合に、当該文字の部分で探索方向が変化して丸角であると誤認されるため、正確に処理できない。特に、文字と罫線は接触するケースが多いため、この誤認は現実には無視できない。第2に、画像が掠れた場合に、当該掠れの部分で探索方向が変化して丸角であると誤認されるため、正確に処理することができない。即ち、本来直線である罫線が部分的に掠れた場合、当該掠れた部分で180°探索の方向が変化するので、丸角であると誤認される。また、この技術によっては点線からなる罫線は抽出できない。

【0009】

以上のように、従来の文字認識装置等の表画像処理装置においては、丸角部分の認識処理の重要性は十分に認識されていながらも、罫線の掠れ、文字の丸い部分の存在、罫線と文字との接触等により、丸角を正確に処理することはできなかった。

【0010】

本発明は、丸角を有する表画像を正確に処理することができる表画像処理装置



を提供することを目的とする。

【0011】

また、本発明は、丸角を有する表面像を正確に処理することができる表面像処理装置を実現するプログラムを記憶したプログラム記憶媒体を提供することを目的とする。

【0012】

【課題を解決するための手段】

図1は本発明の原理構成図であり、本発明による表面像処理装置100の構成を示す。本発明の表面像処理装置100は、入力された画像から縦及び横方向の直線を抽出する直線抽出手段2と、直線抽出手段2の抽出した直線の端点を起点とする斜め成分を抽出し、抽出した斜め成分を用いて丸角候補領域を検出する丸角候補領域検出手段3と、丸角候補領域検出手段3の検出した丸角候補領域を含むセルを抽出するセル抽出手段4と、セル抽出手段4の抽出したセルの角部について丸角部を決定する処理を行う丸角部決定手段5とを備える。

【0013】

本発明の表面像処理装置100によれば、例えば、画像が掠れた場合でも丸角部を正確に処理することができ、縦方向及び横方向のいずれか一方の罫線を抽出できれば丸角部を正確に処理することができる。また、（マッチングパターンを生成させるので）予め多数のマッチングパターンを持つ必要がなく、メモリ容量を大きくする必要がない。更に、罫線に文字が接触する部分や画像の掠れた部分が丸角であると誤認されることがなく、従って、丸角部を正確に処理することができる。このように、本発明の表面像処理装置100においては、罫線の掠れ、文字の丸い部分、罫線と文字との接触等があっても丸角を正確に処理することができ、この結果、罫線構造の認識及びセル（文字認識領域）の抽出を安定して行うことができる。

【0014】

また、本発明によれば、以上に述べた処理を当該表面像処理装置100であるコンピュータに実現させるプログラムを格納したプログラム記憶媒体を提供することができるので、丸角部を正確に処理することができる表面像処理装置100

を容易に提供することができる。

【0015】

【発明の実施の形態】

図2は表画像処理装置構成図であり、本発明の表画像処理装置100の一例の構成を示す。

【0016】

表画像処理装置100は、画像入力装置1、直線抽出装置2、丸角候補領域検出装置3、セル抽出装置4、丸角部決定装置5、文字認識領域決定装置6とを備える。表画像処理装置100は主メモリ上に存在する表画像処理プログラムとCPUとからなる。コンピュータである表画像処理装置100は、図示しないが、表画像処理の対象である画像の入力や指示入力のためのキーボード、マウス、スキャナ等の入力装置、画像の表示等のための表示装置等の出力装置、画像データ12等の各種のデータを記憶するための補助記憶装置を備える。

【0017】

画像入力手段である画像入力装置1は、罫線により構成された帳票の画像である入力画像11を画像データ12として入力する。画像入力装置1は、周知のスキャナ等からなり、帳票等の紙の上に印刷等された縦（垂直又はy）方向及び横（水平又はx）方向の罫線（縦横罫線）及び当該帳票に記入された文字からなる入力画像11を読み取り、電子的なイメージデータである画像データ12として入力する。入力画像11は、特に制限されないが、例えば極端な傾きのない2値画像が好ましく、例えば図18（A）又は図18（B）に示す帳票からなる。

【0018】

直線抽出手段2である直線抽出装置2は、画像入力装置1により入力された画像データ12を参照して、これから縦及び横方向の直線を抽出する。抽出された直線についてのデータ（直線データ）13は、当該直線を表す位置座標からなる。直線データ13は補助記憶装置に記憶される。

【0019】

丸角候補領域検出手段3である丸角候補領域検出装置3は、画像データ12及び直線データ13を参照して、直線抽出装置2の抽出した直線の端点を起点とす

る斜め成分を抽出し、抽出した斜め成分を用いて所定の処理を行うことにより、丸角候補領域を検出する。丸角候補領域検出装置 3 は検出した丸角候補領域を丸角候補データ 14 として補助記憶装置に記憶する。丸角候補データ 14 は、例えば当該斜め成分が丸角候補であるか否かを示すデータと、当該斜め成分を表す位置座標とからなる。

## 【0020】

丸角候補領域検出装置 3 は、前記斜め成分の抽出において、入力された画像データ 12 から直線抽出装置 2 により抽出された縦及び横方向の直線の内、縦方向の直線の端点を起点とする第 1 の斜め成分と、横方向の直線の端点を起点とする第 2 の斜め成分とを別個に抽出する。そして、丸角候補領域検出装置 3 は、丸角候補領域を検出するための前記所定の処理を、以下の第 1 乃至第 3 の処理のいずれかにより行う。即ち、第 1 の処理として、当該第 1 及び第 2 の斜め成分が互いに重なる場合に、当該部分が丸角候補領域であるとする。第 2 の処理として、当該第 1 及び第 2 の斜め成分が互いに予め定められた距離以下の位置関係にあり、かつ、両者の間に直線の特徴を示すパターンがあるか両者が互いに接触又は重なっている場合に、当該部分が丸角候補領域であるとする。第 3 の処理として、当該斜め成分の近傍に他の斜め成分が存在せず、かつ、当該斜め成分の先端に直線の特徴を示すパターンがある場合に、当該部分が丸角候補領域であるとする。

## 【0021】

セル抽出手段 4 であるセル抽出装置 4 は、画像データ 12、直線データ 13 及び丸角候補データ 14 を参照して、丸角候補領域検出装置 3 の検出した丸角候補領域を含むセル（及び他のセル）を抽出する。抽出したセルについてのデータ（セルデータ 15）は、当該セルを表す位置座標からなる。セルデータ 15 は補助記憶装置に記憶される。

## 【0022】

丸角部決定手段 5 である丸角部決定装置 5 は、画像データ 12、丸角候補データ 14 及びセルデータ 15 を参照して、セル抽出装置 4 の抽出した丸角候補領域を含むセルの角部（又は頂点）について所定の処理を行う。これにより、丸角部決定装置 5 は丸角部を決定し、決定した丸角部を丸角部データ 16 として補助記

憶装置に記憶する。なお、実際のセルデータ 15 は丸角候補データ 14 を含むようにされるので、丸角候補データ 14 は参照されず、セルデータ 15 が参照される。

#### 【0023】

丸角部決定装置 5 は、入力された画像データ 12 から抽出された縦及び横方向の直線に基づいてセル抽出装置 4 の抽出したセルの角部について所定の処理を行うことにより、丸角部を決定する。丸角部決定装置 5 は、前記所定の処理を、以下の第 1 乃至第 3 の処理のいずれかにより行う。即ち、第 1 の処理として、セル抽出装置 4 の抽出した当該セルの角部において画素密度が予め定められた順に変化する場合に、当該部分が丸角部であるとする。第 2 の処理として、画素密度の変化に基づいて決定された丸角部が存在する場合に、前記入力された画像の他の角部も丸角部であるとする。第 3 の処理として、当該セルを構成する前記直線抽出手段の抽出した直線の端点と端点とを結ぶように生成された  $n$  次関数からなるパターンと、前記入力された画像とがマッチングする場合に、当該部分が丸角部であるとする。

#### 【0024】

文字認識領域決定手段である文字認識領域決定装置 6 は、画像データ 12、セルデータ 15 及び丸角部データ 16 を参照して、所定の処理を行うことにより、入力された画像データ 12 における文字認識領域を決定する。即ち、丸角部決定装置 5 により決定された丸角部を含むセルについて、当該丸角部を文字領域から除くことにより、当該文字認識領域を決定する。決定した文字認識領域についてのデータ（文字認識領域データ 17）は補助記憶装置に記憶される。なお、実際の丸角部データ 16 はセルデータ 15 を含むようにされるので、セルデータ 15 は参照されず、丸角部データ 16 が参照される。

#### 【0025】

直線抽出装置 2、丸角候補領域検出装置 3、セル抽出装置 4、丸角部決定装置 5、及び、文字認識領域決定装置 6 は、当該表画像処理装置 100 の主メモリ上に存在し CPU 上で実行されることにより各々の処理を実行する処理プログラムによって実現される。直線抽出装置 2 の実行する処理、セル抽出装置 4 の実行す

る処理及び文字認識領域決定装置 6 の実行する処理は周知の処理である。丸角候補領域検出装置 3 の実行する処理及び丸角部決定装置 5 の実行する処理は本発明に特有の処理である。

【0026】

図 3 は表画像処理装置 100 の実行する表画像処理フローを示す。

【0027】

画像入力装置 1 が入力画像 11 を画像データ 12 として入力する画像入力処理を行う（ステップ S1）。即ち、スキャナでの入力画像 11 の読み取りを行い、画像データ 12 を得る。

【0028】

直線抽出装置 2 が画像データ 12 から縦及び横方向の直線を抽出する直線抽出処理を行う（ステップ S2）。即ち、直線データ 13 を得る。この処理については図 4 乃至図 6 を参照して後述する。

【0029】

丸角候補領域検出装置 3 が、直線データ 13 を参照して、抽出された直線の端点を起点とする斜め成分を抽出し、抽出した斜め成分を用いて所定の処理を行うことにより丸角候補領域抽出処理を行う（ステップ S3）。即ち、丸角候補データ 14 を得るこの処理については図 7 乃至図 11 を参照して後述する。

【0030】

セル抽出装置 4 が、画像データ 12、直線データ 13 及び丸角候補データ 14 を参照して、抽出された丸角候補領域を含むセル（及び他のセル）を抽出するセル抽出処理を行う（ステップ S4）。即ち、セルデータ 15 を得る。この処理については図 12 を参照して後述する。

【0031】

丸角部決定装置 5 は、画像データ 12、丸角候補データ 14 及びセルデータ 15 を参照して、抽出された丸角候補領域を含むセルの角部について所定の処理を行うことにより丸角部決定処理を行う（ステップ S5）。即ち、丸角部データ 16 を得る。この処理については図 13 乃至図 16 を参照して後述する。

【0032】

文字認識領域決定装置 6 が、画像データ 12、セルデータ 15 及び丸角部データ 16 を参照して、所定の処理を行うことにより、画像データ 12 における文字認識領域を決定する文字認識領域決定処理を行う（ステップ S6）。即ち、文字認識領域データ 17 を得る。この処理については図 17 を参照して後述する。

#### 【0033】

図 4 は直線抽出装置 2 の実行する直線抽出処理フローを示す。図 4 に示す処理は、図 3 のステップ S2 に相当し、周知の処理である。画像データ 12 から縦及び横方向の罫線（縦横罫線）が抽出される。この直線抽出処理は周知のいずれの処理によっても良いが、以下にその好ましい処理の一例を示す。

#### 【0034】

最初に、連結パターンが抽出される（ステップ S21）。連結パターンは、周知の連結パターン抽出処理を用いることによって抽出される。周知の連結パターン抽出処理としては、例えば、先に本願出願人の提案した特願平 7-282171 号に記載の技術が用いられる。この技術によれば、8 連結で繋がっている画像パターンをラベリングにより抽出する。これにより、複数の枠が配置される位置の相対的な関係に依存することなく、縦横罫線の各パターンを安定に抽出することができる。

#### 【0035】

ここで、当該連結パターン抽出処理について簡単に説明する。まず、入力された画像データ 12 を構成するパターン（部分パターン）の各々について、その縦（上下の 2 方向）、横（左右の 2 方向）及び斜め（これらの中間の 4 方向）の合計 8 方向において、他の部分パターンとの連結を調べる。そして、いずれかの方向において連結している部分パターン（連結パターン）に対して同一のラベルを貼付する。連結パターンは、文字が接触していない枠即ち罫線（又は罫線の一部）、文字が接触している罫線（又は罫線の一部）、罫線に接触していない文字（又は文字の一部）のいずれかである。

#### 【0036】

次に、ラベリングされた部分パターンを判別し、所定のサイズより大きい連結パターンの集合を当該帳票を構成する罫線の候補として抽出する。即ち、抽出さ

れた部分パターンの内、一定以上の大きさを有する部分パターン（第1の部分パターン）を帳票の罫線の候補とする。例えば、本来の罫線の内、掠れることなく正しく抽出された罫線が第1の部分パターンとされ、一方、本来の文字の内、罫線に接触することなく正しく記入された文字は罫線の候補から除外される。

## 【0037】

次に、第1の部分パターンに対して所定の位置に存在し、かつ、所定のサイズより大きな部分パターン（第2の部分パターン）に貼付されたラベルを、当該第1の部分パターンのそれと同一のラベルに変更することにより、当該第2の部分パターンを帳票の罫線の候補に追加する。例えば、本来の罫線の内、掠れてしまって第1の部分パターンから孤立した罫線が第2の部分パターンとされる。この結果、掠れた罫線も第1の部分パターン即ち罫線の候補とされ、正しい結果が得られる。従って、このような連結パターン抽出処理を用いることによって、本来罫線を構成するパターンが掠れにより分離して別のパターンとして処理されることを防止することができる。

## 【0038】

この後、抽出した連結パターンの画像について、当該連結パターン（部分パターン）毎に、縦横双方向の成分の細分化（マスク処理）が実行される（ステップS22）。縦横双方向の成分の細分化は、周知のマスク処理を用いることによって行われる。周知のマスク処理としては、例えば、先に本願出願人の提案した特開平7-28937号公報に記載の技術が用いられる。この技術によれば、抽出された連結パターンの画像から極端な斜め成分を取り除く。これにより、枠即ち罫線だけに存在する長い直線を抽出し易くすることができる。

## 【0039】

ここで、当該マスク処理について簡単に説明する。まず、ラベリングにより得た部分パターン（連結パターン）毎に、所定の大きさ（長さ及び幅）を有する横長及び縦長の2種類のマスクを用いた走査を行う。そして、各マスク毎に、当該マスクの面積に対するマスク内のパターンの占める面積の割合を算出し、当該割合が所定値よりも大きければそのマスク内を全てパターンとみなす（画像の横成分又は縦成分とする）。即ち、当該部分を当該マスクの大きさに細分化された矩

形領域とみなす。当該割合が所定値以下であればそのマスク内のパターンを削除する（画像の横成分又は縦成分が存在しないとする）。これにより、画像データ 12 を構成する画像パターンの縦横成分を、使用したマスクの幅の程度にまで細分化して抽出することができる。

#### 【0040】

次に、複数の行（横方向）又は列（縦方向）が連続して上記割合が所定値よりも大きくなる場合、それらの行又は列を統合して 1 個の矩形の領域とし、当該矩形範囲の中心線（図 5 の左側の直線）を処理結果とする。従って、このようなマスク処理を用いることによって、枠即ち罫線の太さを同一にして、罫線のみが存在する長い直線を抽出し易くすることができる。

#### 【0041】

この後、マスク処理した画像について、線分抽出処理が実行される（ステップ S23）。線分抽出は、周知の線分抽出処理を用いることによって行われる。周知の線分抽出処理としては、例えば、先に本願出願人の提案した特開平 6-309498 号公報に記載の技術が用いられる。この技術によれば、マスク処理した画像に対して隣接投影値を算出し、縦横双方向について、矩形近似により直線又は直線の一部（線分）を検出する。ここで、隣接投影とは、注目するある行又は列の投影値に、その周囲の行又は列の投影値を足し合わせ、この値を投影値とする処理を言う。線分抽出処理の結果を図 5 に示す。これにより、画像の着目点の周囲を大局的に捉え線分を正確に抽出することができる。

#### 【0042】

ここで、当該線分抽出処理について簡単に説明する。まず、マスク処理の結果である（当該マスク内を全てパターンとされた）矩形範囲のパターンの各々について、その横方向の投影と縦方向の投影をとり、投影値を求める。そして、求めた投影値に対してその周囲（注目するある行又は列）のパターンについて求めた投影値を所定の範囲において加算する。即ち、特開平 6-309498 号公報に記載の式（例えば（11）式）に従って、当該パターンについての行及び列方向の隣接投影値を算出する。

#### 【0043】



次に、行方向の算出した隣接投影値と矩形範囲（マスク）の横方向の分割長との比及び列方向の算出した隣接投影値と矩形範囲（マスク）の縦方向の分割長との比を算出する。そして、これらの比が所定の閾値以上である部分を直線の候補とし、当該比が閾値よりも小さい部分を線分が存在しないものとする。そして、複数の部分が連続して所定の閾値以上となる場合、即ち、直線の候補である部分が連続する場合、これらを統合して1個の矩形領域で近似し、この近似された矩形範囲に直線が存在するものとする。即ち、図5の右側に示すように、矩形領域の連続により近似された線分が抽出される。

#### 【0044】

このような線分抽出処理を用いることによって、直線が傾きによって斜め方向に複数行又は列にわたっていても、線分を抽出することができる。従って、大きな枠即ち罫線についても、それが傾いていても、その線分を正確に抽出することができる。また、隣接投影を用いることにより、通常の投影のようにマスクの分割長を短くして分割数を大きくすることにより傾いた直線を抽出する必要がないので、文字を構成する短い線分を抽出することなく、長い線分を抽出することができる。

#### 【0045】

この後、線分抽出処理の結果を用いて、直線抽出処理が実行される（ステップS24）。直線抽出は、周知の直線抽出処理を用いることによって行われる。周知の直線抽出処理としては、例えば、先に本願出願人の提案した特開平6-309498号公報に記載の技術が用いられる。この技術によれば、線分抽出処理において抽出され矩形近似された線分の内、近隣の線分どうしを統合して長い直線を検出し、検出した直線に対して更に矩形近似を行う。直線抽出処理の結果を図6に示す。これにより、直線を忠実に抽出することができる。

#### 【0046】

ここで、当該直線抽出処理について簡単に説明する。まず、線分抽出処理において抽出され矩形近似された線分（図6の左側に示す線分）がその上下（例えば、隣接投影値の算出において加算する範囲）に存在する他の線分と接触しているかを調べる。次に、接触する他の線分が存在する場合、当該線分とその上下に存

在する他の線分とを統合して、図6の右側に示す1個の矩形領域からなる線分（矩形線分）とする。このような直線抽出処理を用いることによって、帳票の罫線等の長い直線を抽出することができる。

【0047】

なお、特開平6-309498号公報においては、隣接投影値と矩形範囲（マスク）分割長との比の算出処理以降を、線分抽出処理においてではなく、直線抽出処理において行っている。しかし、本願においても上述の説明のように線分抽出処理（ステップS23）と直線抽出処理（ステップS24）とは連続して行われるので、実際には、特開平6-309498号公報における処理と何ら相違ない処理が行われる。

【0048】

この後、線分抽出処理の結果を用いて、画像データ12からの点線抽出処理が実行される（ステップS25）。点線抽出は、周知の点線抽出処理を用いることによって行われる。周知の点線抽出処理としては、例えば、先に本願出願人の提案した特開平9-50527号公報に記載の技術が用いられる。この技術によれば、同一サイズの部分パターンが規則正しく並んでいる部分を点線として抽出し、直線と同様に矩形近似する。なお、抽出された直線（実線）及び点線は、区別されることなく、罫線として同様に処理される。これにより、点線からなる罫線を含む帳票を以後の処理において正確に処理することができる。

【0049】

ここで、当該点線抽出処理について簡単に説明する。まず、連結パターン抽出処理（ステップS21）において抽出しラベリングしたパターンの最も外側の矩形の画像の一部を当該画像を縦又は横方向に横切るような第1の短冊として取り出す。そして、その内部に存在するパターン（内部パターン）を予め定められた点線を構成する所定の大きさのパターン（予定パターン）と比較し、予定パターンに一致する部分パターンを探す。

【0050】

次に、予定パターンと一致する内部パターンが見つかった部分を、更に、当該短冊を垂直な方向に横切るような第2の短冊として取り出す。そして、その内部

において、予定パターンと一致する部分パターンを抽出する。

【0051】

次に、抽出した部分パターンがある一定の間隔で並びかつある一定の数以上連続して並んでいる場合、当該並んでいるものを点線と判断して、これを囲む矩形領域を点線の存在する範囲として、当該矩形により近似する。

【0052】

このような点線抽出処理を繰り返し実行することにより、帳票内に存在する点線を正確に抽出して矩形領域により近似することができる。従って、以後野において、実線及び点線を同様なものとして処理することができる。

【0053】

以上の処理の結果、画像データ12から図6の右側に示すように矩形により近似された直線及び点線が抽出される。これが直線データ13である。

【0054】

図7は丸角候補領域検出装置3の実行する丸角候補領域検出処理フローを示す。図7に示す処理は、図3のステップS3に相当し、本発明に特徴的な処理である。丸角候補領域検出処理は、前処理としての斜め成分抽出処理と、実際の丸角候補領域を検出する処理である候補検出処理とからなる。

【0055】

最初に、斜め成分抽出処理が実行される（ステップS31）。斜め成分（斜め線）の抽出は直線抽出装置2において抽出された縦横の直線の端部を起点として行われる。これにより、丸角部の候補の位置を正確に検出することができる。斜め成分抽出処理には、斜め方向の画素探索処理又は斜め方向の投影処理等の処理が用いられる。

【0056】

斜め方向の画素探索は、周知の斜め方向の画素探索処理を用いることによって行われる。周知の斜め方向の画素探索処理としては、例えば、先に本願出願人の提案した特開平7-192094号公報に記載の技術が用いられる。この技術によれば、抽出された直線の端点から画素の存在（パターンの輪郭）を探索することにより、斜め線を抽出する。これにより、斜め線に多少の掠れがあっても斜め

成分を抽出することができる。

【0057】

ここで、当該斜め方向の画素探索処理について簡単に説明する。まず、直線抽出装置2において抽出された縦横の直線の端部を起点（開始点）として定める。そして、当該直線を除く当該開始点から始まるパターンの輪郭を探索する。即ち、当該直線の下辺あるいは上辺又は左辺あるいは右辺から、当該パターンを構成する画素の存在する線に沿って辿る。そして、当該直線又は当該直線以外の他の縦又は横の直線に達したら探索を終了し、当該探索の結果を斜め成分とする。

【0058】

また、斜め方向の投影は、周知の斜め方向の隣接投影処理を用いることによって行われる。周知の斜め方向の隣接投影処理としても、例えば、先に本願出願人の提案した特開平7-192094号公報に記載の技術が用いられる。この技術によれば、画像データ12の画像パターンに対して斜め方向の隣接投影値を算出し、当該斜め方向について、矩形近似により斜め方向の線又は線の一部（線分）を検出する。ここで、隣接投影とは、注目する斜め線の投影値に、その周囲の投影値を足し合わせ、この値を投影値とする処理を言う。これにより、斜め線に多少の掠れがあっても斜め成分を抽出することができる。

【0059】

ここで、当該斜め方向の投影処理について簡単に説明する。まず、画像データ12の画像パターンの内、直線抽出処理（ステップS2）において直線として抽出されたパターン（直線データ13に対応するパターン）を除いた各々について、表す定められた矩形範囲に分割する。そして、分割した各々の矩形範囲についてその縦及び横方向の投影をとり、投影値を求める。そして、求めた投影値に対してその周囲の矩形範囲の縦及び横方向について求めた投影値を所定の範囲で加算する。即ち、特開平7-192094号公報に記載の式に従って、当該矩形範囲についての斜め方向の隣接投影値（斜め隣接投影値）を算出する。投影の方向は、例えば45°の角度である。

【0060】

次に、算出した斜め隣接投影値と矩形範囲の分割長との比を算出する。そして

、この比が所定の閾値以上である部分を斜め成分とし、当該比が閾値よりも小さい部分を斜め成分が存在しないものとする。そして、複数の部分が連続して所定の閾値以上となる場合、即ち、斜め成分である部分が連続する場合、これらを統合して1個の矩形領域で近似し、この近似された矩形範囲に斜め成分が存在するものとする。斜め成分は複数の矩形範囲からなる。

## 【0061】

このような斜め方向の画素探索処理又は斜め方向の（隣接）投影処理を用いることによって、斜め方向に傾いている斜め成分を正確に抽出することができる。従って、帳票の枠における丸角部を構成する斜め成分を正確に抽出することができる。

## 【0062】

斜め成分抽出処理により抽出された斜め成分の内、第1の斜め成分である縦方向の直線の端部を起点とするものを「斜め成分T」と呼び、第2の斜め成分である横方向の直線の端部を起点とするものを「斜め成分Y」と呼んで、区別することとする。

## 【0063】

図8に丸角部分を拡大した例を示す。直線抽出装置2において縦横罫線を抽出すると、図8（A）において点線で囲んだ部分が縦直線及び横直線（直線データ13）として抽出される。そこで、図8（B）に示すように、横直線の端部から斜め方向に伸びているパターンを抽出し、その結果として、当該斜め成分Y及びその存在する領域（斜線を施した領域）を記憶する。次に、図8（C）に示すように、縦直線の端部から斜め方向に伸びているパターンを抽出し、その結果として、当該斜め成分T及びその存在する領域（斜線を施した領域）を記憶する。なお、抽出の順序はいずれが先であっても良い。この処理は全ての縦直線及び横直線について実行される。即ち、本来1個の斜め成分が斜め成分T及びYとして別々に抽出される。これにより、1個の斜め成分が完全に掠れていても斜め成分T及びYとして抽出することができる。

## 【0064】

斜め成分抽出処理に続いて、候補検出処理が実行される（ステップS32）。

候補検出処理は、重なり処理、近接処理又は探索処理の内の1又は複数の処理からなる。重なり処理については図8を参照して説明し、近接処理及び探索処理の例については、各々、図9及び図10に示す。

#### 【0065】

最初に、第1の処理として、重なり処理が実行される（ステップS32-1）。重なり処理は、例えば図8（B）に示す斜め成分Y（の存在する領域）と図8（C）に示す斜め成分T（の存在する領域）とが重なっている場合に、実行される。従って、ステップS32-1において最初に、斜め成分Yと斜め成分Tとの重なりの有無が調べられる。当該重なりが無い場合、ステップS32-1における以下の処理は省略される。斜め成分Yと斜め成分Tとが重なっている場合、その部分は丸角候補領域であるとされる。これは、両者が重なる場合には、図8に示すように、当該斜め成分が丸角部の備えるべき形状を備えることになるからである。即ち、当該斜め成分は、横直線及び縦直線の双方に接続しており、かつ、当該丸角部は双方の直線を繋ぐ斜め成分である。この重なり処理により丸角部を正確に抽出することができる。相互に重なる斜め成分Y及びTが無い場合、図9に示す近接処理が実行される。

#### 【0066】

なお、当該斜め成分Y及びTは完全に重なっている必要はなく、予め定められた状態以上に双方が重なっていれば、両者が重なっていると判断される。重なり状態は、例えば重なっている面積又は画素数又はその割合により判断される。これにより、掠れがあっても斜め成分を抽出することができる。

#### 【0067】

次に、第2の処理として、近接処理が実行される（ステップS32-2）。近接処理は、2つの斜め成分Y（の存在する2つの領域）が縦方向に、又は、斜め成分T（の存在する2つの領域）が横方向に、相互に予め定められた距離以下の位置関係にある場合に、実行される。従って、ステップS32-2において最初に、処理対象として着目したある斜め成分Y又はTの近傍の他の斜め成分Y又はTとの距離が調べられる。前記位置関係を満たす他の斜め成分Y又はTが有る場合、図9に示すように、2つの領域Y又はTの間の距離に応じて、更に、2つの

異なる処理のいずれか一方が実行される。2つの異なる処理はいずれを先に実行しても良い。前記位置関係を満たす他の斜め成分Y又はTが無い場合、図10に示す探索処理が実行される。

## 【0068】

なお、図9は2つの斜め成分Yが縦方向に相互に前記位置関係にある場合について示すが、2つの斜め成分Tが横方向に相互に一定距離以下の位置関係にある場合についても同様である。また、図10は斜め成分Yについての探索処理を示すが、斜め成分Tについての探索処理も同様である。

## 【0069】

図9(A)に示すように、2つの斜め成分Yの間が予め定められた距離以下ではあるが上下方向に離間している場合、当該2つの斜め成分Yの間に存在するパターンを調べる。当該パターンが直線の特徴を示すパターンでない場合、以下の処理は省略される。直線の特徴を示すパターンである場合、当該2つの斜め成分Yの間に仮想的に直線を想定し、当該2つの斜め成分Yを丸角候補領域であるとする。仮想的な直線は罫線候補領域であるとされる。そして、罫線候補領域内に存在する画像のパターンSが直線の特徴を示すパターンであるか否かが調査される。直線の特徴を示す場合、当該領域は罫線として抽出され、両端の斜め成分Yは丸角候補領域として抽出される。これにより、文字の一部を丸角部として誤って抽出することを防止することができる。従って、2つの斜め成分Yが上下方向に（又は2つの斜め成分Tが左右方向に）一定距離以下の位置関係にある場合には、これらの間に直線成分が殆ど存在せず略円弧状になっていても、丸角候補領域を正確に抽出することができる。

## 【0070】

当該パターンSが直線の特徴を示すパターンであるか否かを調べる処理（直線探索処理）は、周知の直線探索処理を用いることによって実行される。周知の直線探索処理としては、例えば、先に本願出願人の提案した特願平9-141470号に記載の技術が用いられる。この技術によれば、画素の探索及び凹凸度による直線判定が実行される。これにより、凹凸度の小さい直線と凹凸度の大きい文字や斜め成分とを区別することが可能になり、直線の特徴を示すパターンである

か否かを正確に調べることができる。

【0071】

ここで、当該画素の探索及び凹凸度による直線判定処理について簡単に説明する。まず、入力された画像データ12における当該調査対象の画像パターンSの延びる方向（縦又は横方向）に、画素（パターンとして入力された黒画素）を探索する。そして、当該探索方向に画素が存在する限り、当該探索方向に探索を続ける。この場合、画素の探索方向は変化しない。一方、当該探索方向に画素が存在しない場合に限り、当該探索方向の斜め方向にある画素を探索する。この場合、画素の探索方向が変化する。当該探索方向の変化回数がカウントされる。

【0072】

次に、画素の探索方向の変化回数の値に基づいて、当該パターンSが直線の特徴を示すパターンであるか否かが判定される。当該パターンSが直線であれば略同一方向に探索が行われるので画素の探索方向の変化回数の値が小さく、直線でなければ当該値が大きい。そこで、予め経験的に直線判定の閾値を適切な値に定めることにより、これを用いて直線とするか否かを判定する。このような画素の探索及び凹凸度による直線判定処理を用いることによって、対象とするパターンSが直線か否かを正確に判定することができる。

【0073】

なお、図9（A）に示す場合において、当該2つの領域Yの間に存在するパターンSについて直線探索処理を行うことなく、丸角部が接近して存在する領域であると判断しても良い。この場合、2つの領域Yの間の直線（罫線）は、上述のようにして作成しても良く、又は、当該2つの領域Yの間におけるパターンの存在の有無にかかわらず作成しても良い。

【0074】

図9（B）に示すように、2つの領域Yの間が接触しているか又は重なっている場合、当該2つの領域Yが丸角候補領域であるとされる。ここで、接触とは2つの斜め成分Yを構成する画素が隣接している場合を言い、重なるとは2つの斜め成分Yを構成する画素が重複している場合を言う。このような状態は、2本の横直線に挟まれた縦直線の成分が短すぎるため、直線として抽出されなかった場



合である。従って、この処理は、2つの領域Yの間の間隔が直線として抽出される最小長さ以下である場合に実行される。これにより、2つの斜め成分Yが上下方向に又は2つの斜め成分Tが左右方向に接触又は重なる場合には、これらの間に直線成分が殆ど存在せず略円弧状になっていても、丸角候補領域を正確に抽出することができる。

## 【0075】

次に、第3の処理として、探索処理が実行される（ステップS32-3）。探索処理は、図10に示すように、ある斜め成分Y（の存在する領域）又は斜め成分T（の存在する領域）の周囲に、当該斜め成分と予め定められた距離以下の位置関係にある他の斜め成分Y又はTが存在しない場合に、実行される。

## 【0076】

図10に示すように、抽出されたある斜め成分（Y1）の領域から一定距離以下の位置に他の斜め成分Y又はTが存在しない場合、当該斜め成分（Y1）の先端に短い直線成分が存在するか否かが調べられる。具体的には、斜め成分Yの場合には縦直線が探索され、斜め成分Tの場合には横直線が探索される。ここで、当該短い直線の存在を調べる処理（短い直線探索処理）は、周知の短い直線探索処理を用いることによって行われる。周知の短い直線探索処理としては、例えば、先に本願出願人の提案した特願平9-141470号に記載の技術が用いられる。この技術によれば、画素の探索及び凹凸度による直線判定処理が実行される。この処理については、「直線探索処理」として前述したので、その説明を省略する。

## 【0077】

当該直線が存在する場合、当該部分に直線を想定し、当該領域Y1を丸角候補領域であるとする。当該直線を想定することにより、当該直線は直線抽出処理において抽出した直線と同様のもの（罫線）であるとされる。これにより、抽出された斜め成分の近傍に他の斜め成分が存在しない場合でも、直線の特徴を示すパターンS'が存在していれば、当該抽出された斜め成分は丸角候補領域であることができる。従って、横又は縦方向に直線成分が殆ど存在せず帳票の角部が円弧状になっている場合でも、丸角候補領域を正確に抽出することができる。

また、短い罫線でも正確に抽出することができる。

【0078】

なお、図10に示す斜め成分Y2は図8に示す場合に相当し、対比のために示される。即ち、斜め成分Y1は、前述の重なり処理及び近接処理によっては、当該斜め成分Y1が丸角候補領域であるということを抽出できない場合である。

【0079】

次に、以上の各種の処理により丸角候補領域として抽出された斜め成分について、直線情報の作成処理が実行される（ステップS33）。即ち、当該斜め成分を、縦方向及び横方向に分解し、各々の成分を縦方向及び横方向の仮想的な直線（罫線）とする。具体的には、図11に示すように、前記斜め成分の横線に接する端点から、当該縦方向の座標値をそのままにして、前記斜め成分の接する縦線の持つ横方向の座標まで、当該横方向の座標値を変化させることにより、横直線を得る。また、同様にして、縦直線を得る。そして、これらの2本の直線の存在を示す2つの矩形領域（図11において点線で示す）が作成される。これにより、仮想的な直線（矩形領域）についての直線情報が作成される。従って、直線情報は当該仮想的な直線の始点及び終点（矩形領域の範囲）を示す位置座標からなる。この後、仮想的な直線と図11において実線で示す実在する直線（矩形領域）とを統合し、その結果を用いて以下のセル抽出が実行される。これにより、丸角候補領域として抽出された斜め成分を仮想的な直線として処理できるので、丸角部を意識することなく、セル抽出を実行することができる。

【0080】

なお、仮想的な直線（罫線）と実際に抽出された罫線とは、完全に接触又は重なっていなくても予め定められた一定距離以下の位置関係にある場合、統合されて1本の直線（罫線）とされる。これは縦及び横の双方について同様である。従って、この統合処理以降の処理においては、実在の罫線と仮想の罫線とが同様に扱われることにより、丸角部を意識することなく、セル抽出を実行することができる。

【0081】

また、この直線情報作成処理と共に、丸角候補領域であるとして抽出された斜

め成分のサイズの平均値を算出する処理が実行される。この平均値は後述する丸角部決定処理において用いられる。

## 【0082】

また、直線情報作成処理は以下に述べるセル抽出処理の前処理であるので、当該セル抽出処理に含めても良い。即ち、セル抽出装置4が実行するようにしても良い。

## 【0083】

以上の処理の結果、図11に示すように、直線データ13を用いることにより、画像データ12から帳票の角部における丸角候補領域が抽出され、これについての直線情報が作成される。これが丸角候補データ14である。

## 【0084】

図12はセル抽出装置4の実行するセル抽出処理について説明する。図12に示す処理は、図3のステップS4に相当し、周知の処理である。

## 【0085】

セル抽出装置4は画像データ12からセルを抽出する。この抽出処理は周知のどのような処理によっても良い。セルは四方を直線で囲まれた領域であるので、2本の横線の組（横枠）と2本の縦線の組（縦枠）とでセルを規定（抽出）することができる。以下にその一例を示す。

## 【0086】

最初に、横枠を決定する横枠決定処理が実行される（ステップS41）。横枠決定は、周知の横枠決定処理を用いることによって行われる。周知の横枠決定処理としては、例えば、先に本願出願人の提案した特開平9-50527号公報に記載の技術が用いられる。この技術によれば、即ち、枠の上部から順番に、所定の規則に従って、一行を構成する2本の横線の組（対）を決定し、これを横枠とする。

## 【0087】

ここで、当該横枠決定処理について簡単に説明する。まず、直線抽出処理（ステップS2）において抽出された横線の中から、上から順番に2本の横線を選択して、これらを2本の横線の組の候補とする。

## 【0088】

次に、この候補である2本の横線が、同一の長さであるか、又は、下方の直線の方が長いという関係（組の条件）であれば、当該候補をそのまま2本の横線の組とする。下方の直線の方が長い場合、当該下方の直線は、再度、2本の横線の組の候補とすることができる。

## 【0089】

一方、候補である2本の横線が、下方の直線の方が短いという関係であれば、更に、当該下方の直線とその下方の直線とを新たな2本の横線の組の候補とする。そして、双方の直線の長さの関係を調べ、上述の組の条件を満たし当該候補を2本の横線の組とすることができるまで、同様の処理を繰り返す。なお、下方の直線の更に下方に上述の組の条件を満たす直線が全く存在しない場合、当該下方の直線とその直ぐ上の直線とを2本の横線の組とする。

## 【0090】

次に、上述の処理を最も下の直線について行った後、当該直線の上方に未処理の直線が残っている場合、当該未処理の直線と再度2本の横線の組の候補とすることができる直線とを用いて、再度、上述の処理を繰り返すことにより、2本の横線の組を作る。このような横枠決定処理を用いることによって、以後の処理において行等の矩形枠を構成する横枠の組（2本の横線の組）を正確に決定することができる。

## 【0091】

この後、横枠決定処理の結果を用いて、縦枠を決定する縦枠決定処理が実行される（ステップS42）。縦枠決定は、周知の縦枠決定処理を用いることによって行われる。周知の縦枠決定処理としては、例えば、先に本願出願人の提案した特開平9-50527号公報に記載の技術が用いられる。この技術によれば、横枠決定処理において組とされた2本の横線により構成される間隔毎に（即ち、一行毎に）処理を行ない、先に抽出された縦線の中で当該2本の横線の双方について上下共に達しているものを縦枠とする。このような縦枠抽出処理を用いることによって、縦枠を正確に抽出することができる。

## 【0092】

この後、横枠決定処理及び縦枠決定処理の結果を用いて、セルである矩形領域を抽出する矩形抽出処理が実行される（ステップS43）。矩形抽出は、周知の矩形抽出処理を用いることによって行われる。周知の矩形抽出処理としては、例えば、先に本願出願人の提案した特開平9-50527号公報に記載の技術が用いられる。この特開平9-50527号公報に記載の処理技術によれば、横枠決定処理及び縦枠決定処理において決定された横枠及び縦枠に四方を囲まれた範囲を、矩形範囲として抽出する。即ち、横枠決定処理において組とされた2本の横線と、縦枠抽出処理において当該横枠に達する縦枠とされた2本の縦線とで囲まれた領域を、矩形枠（セル）として抽出する。このような矩形抽出処理を用いることによって、セルを正確に抽出することができる。

## 【0093】

以上の処理の結果、直線データ13を用いることにより、画像データ12から帳票におけるセルが抽出される。これがセルデータ15である。この時、ステップS33において作成された直線情報により表される仮想の直線を用いることにより、図11に示すような帳票の角部における丸角部を含むセルも、同様に、矩形枠として抽出される。

## 【0094】

図13は丸角部決定装置5の実行する丸角部決定処理フローを示す。図13に示す処理は、図3のステップS5に相当し、本発明に特徴的な処理である。丸角部決定処理は、画素密度変化処理、対象性処理及びマッチング処理からなる。

## 【0095】

最初に、第1の処理として、抽出されたセルの角部の画素密度変化に基づいて当該角部の形状を決定する画素密度変化処理が実行される（ステップS51）。画素密度変化処理は、セルの角部に注目し、当該角部の画素密度の変化を用いて丸角部の決定を行う処理である。即ち、帳票の角部の形状に基づいて丸角部を決定する処理モードである。この処理により、角部に掠れがある場合でも丸角部を決定することができる。ここで、 $(\text{画素密度}) = (\text{当該領域内の黒画素の総数}) / (\text{当該領域内の画素の総数})$ である。黒画素はステップS31において当該斜め成分Y又はTとして抽出された線分を構成する画素である。

## 【0096】

図14に示すように、セルの角部においては、その外側から内側へ向かって、画素密度が低い三角領域a、高い台形領域（斜線を施した部分）b、低い台形領域cの順に並ぶ。即ち、画素密度は、「低い→高い→低い」の順に変化する。そこで、抽出されたセルを構成する2本の直線に挟まれる角部において、当該2本の直線の作る頂点から離れる方向へ上述の順に画素密度が変化している場合、当該部分は丸角部であると決定する。この画素密度が「低い→高い→低い」の順に変化するという特徴を用いることにより、角部に掠れがあっても丸角部を決定することができる。

## 【0097】

画素密度の高低を判断するための閾値（又はパラメータ）が、例えば当該帳票（入力画像11）の入力時にユーザにより、定められる。ユーザの入力がない場合、予め定められた値とされる。図14から判るように、領域aの画素密度は極めて低い（この例では「0」である）。従って、領域aについての閾値は極めて小さくされ、画素密度が当該閾値よりも小さい場合に画素密度が「低い」と判断される。領域bの画素密度は所定のレベルに達する。この所定のレベルは経験的にある程度正確に知り得る。従って、領域bについての閾値は当該レベルより小さい値とされ、画素密度が当該閾値よりも大きい場合に画素密度が「高い」と判断される。領域cの画素密度は、領域aよりは高いが、領域bよりはかなり低い。この画素密度は経験的にある程度正確に知り得る。従って、領域cについての閾値は、領域aについての閾値よりは高いが、領域bよりはかなり小さくされ、画素密度が当該閾値よりも小さい場合に画素密度が「低い」と判断される。

## 【0098】

画素密度処理において、画素密度を算出する領域は、丸角候補領域検出装置3において算出された丸角候補領域として抽出された斜め成分のサイズの平均値を規準として設定される。例えば、当該平均値より所定の割合だけ大きい値が、前記台形領域cの下辺の値とされる。これにより、画素密度変化処理の領域を小さくすることができるので、画素密度変化処理に要する時間を短縮することができる。台形領域cの下辺は当該矩形領域を定める縦及び横の直線に対して、図14

に示すように45°で交わる。台形領域bの上辺及び下辺も、同様である。

【0099】

なお、画素密度処理において、セルを構成する2本の直線の作る頂点を含む三角領域aの画素密度が当該閾値よりも小さい場合に、当該角部が丸角であると判断しても良い。また、領域aの画素密度が領域bの画素密度よりも十分に小さい場合（又は、両者の間に予め定められた差がある場合）に、当該角部が丸角であると判断しても良い。また、領域a、b及びcの画素密度が「低い→高い→低い」と変化し、領域aの画素密度が領域cの画素密度よりも小さい場合に、当該角部が丸角であると判断しても良い。

【0100】

この後、第2の処理として、対象性処理が実行される（ステップS52）。入力された画面データ12が常に良好な状態であるとは限らない。即ち、罫線の掠れや潰れ、罫線への文字の接触等の原因により、画素密度変化処理のみで全ての丸角部を正確に抽出できるとは限らない。そこで、丸角部抽出の対象である表構造の規則性が予め判っている場合、当該規則性を利用することにより、より高い精度で丸角部の抽出を行う。

【0101】

対象性処理は、帳票における表構造の規則性に注目し、角部の同一形状の規則的な繰り返し（丸角部の方向は異なる）を利用して丸角部の決定を行う処理である。一般に用いられる帳票の多くは、図15（A）に示すようにその外周の角部の全てが丸角であるか、逆に、その外周の角部の全てに全く丸角がないかの何れかである。図15（B）に示すように、その外周の角部の一部のみが丸角である帳票は殆ど存在しない。そこで、この対象性を利用して、より高い精度で丸角部を抽出する。

【0102】

最初に、前処理として当該規則性が予め判っているか否かが調べられる（ステップS52-1）。当該帳票が規則性を持った表構造を有することは、例えば当該帳票（入力画像11）の入力時にユーザにより入力される。

【0103】

この後、図15(A)に示す帳票について規則性を有することが指示されている場合、当該規則性(対象性)を利用した処理を行う(ステップS52-2)。即ち、例えば、4個の角部がある当該帳票において四隅の内の何処か1個でも先の画素密度変化処理において丸角部であると判断されると、他の3個の角部も丸角部である可能性が高いと判断する。そして、当該他の3個の角部について、画素密度変化処理の制限を緩めた上で、再度、画素密度変化処理を行い、この結果を最終的な処理結果とする。従って、対象性処理は、実際は、緩和された一定条件下での画素密度変化処理の繰り返しである。上記制限の緩和は、例えば閾値(又はパラメータ)を適切に変更すること等による。

## 【0104】

例えば、図14に示した丸角部において、これに交差するように文字が記入されていた場合、領域a及び領域cの画素密度は当該閾値よりも大きくなる。この場合、画素密度は領域a、b及びcについて高いままで変化しない。また、当該丸角部の極めて近傍まで文字が記入されていた場合、領域cの画素密度が当該閾値よりも大きくなる。この場合、画素密度は領域a、b及びcについて「低い→高い→高い」となる。そこで、上記対象性処理に基づいて領域a及び領域cの閾値を予め定められた値だけ高くする。これにより、上記のような場合でも、画素密度処理における画素密度が領域a、b及びcについて「低い→高い→低い」と変化することになる。結果として、丸角部の決定の緩和が図られ、最初の処理では決定から漏れた角部も丸角部であると決定することができる。

## 【0105】

なお、規則性を有することが指示されている場合であっても、先の画素密度変化処理において予め定められた個数の丸角部が決定されていなければ、対象性処理(ステップS52-2)は実行されない。この場合は帳票の外周の角部の全てに全く丸角がないと考えられるからである。対象性処理を開始するための丸角部の個数は予め定められる。この個数は、前述のように、通常1個とされるが、2個以上であっても良く、例えば当該帳票(入力画像11)の入力時にユーザにより入力される。

## 【0106】



また、対象性処理（ステップ S 5 2-2）は複数回繰り返して実行しても良い。即ち、領域 a 及び領域 c の閾値を所定幅だけ大きくしても丸角部が決定できない場合、更に、領域 a 及び領域 c の閾値を更に同じ所定幅だけ大きくして画素密度変化処理を行っても良く、これを繰り返しても良い。これにより、丸角部に文字が接触等していても確実に丸角部を決定することができる。

## 【0107】

以上のように、帳票の角部の形状に基づいて丸角部を決定する処理モード以外に、帳票の外周における角部の形状の規則性を用いて（及び帳票の角部の形状に基づいて）丸角部を決定する処理モードにより、角部毎に丸角か否かが決定される。これにより、予め帳票の規則性が判っている場合には、より高精度で丸角部を決定することができる。即ち、画素密度変化処理において丸角部が予め定められた個数以上決定された場合には、その後の規則性を用いた処理モードにおいて、画素密度変化処理において丸角部であることを決定できなかった角部についても再度画素密度変化処理を行う。これにより、確実に丸角部であることを決定することができる。

## 【0108】

なお、ステップ S 5 2-1 において処理対象の帳票について予め規則性が判らない場合には、ステップ S 5 2-2 の実行は省略される。即ち、画素密度変化処理において抽出した丸角部のみを丸角部であるとする。従って、この場合、図 15（B）に示すように、その外周の角部の一部のみが丸角である表構造が検出される。

## 【0109】

また、当該帳票が規則性を有するか否かの指示入力を、例えば規則性を有する角部の位置又は数を指定して入力するようにしても良く、また、規則性を有する角部は何箇所でも良い。

## 【0110】

この後、第 3 の処理として、マッチング処理が実行される（ステップ S 5 3）。画素密度変化処理及び対象性処理によっては、点線の内、点と点の間隔が広いものについては処理できない。これは、当該間隔の広い部分の存在により画素密

度の変化が正確に抽出できないためである。そこで、画素密度変化処理（又は画素密度変化処理及び対象性処理の双方）において丸角部であると決定されなかった角部について、マッチング処理が実行される。マッチング処理は、発生させたパターンと画像データ 12 上の画像パターン（又は近似させた画像パターン）とのマッチングを採ることにより、角部の形状を認識する処理である。マッチング処理を行うことにより、点と点の間隔が広い点線についても丸角部を決定することができる。

## 【0111】

マッチング処理は、図 16 に示すように、横線と横線（縦線と縦線でも同様である）及び縦線と横線との間の距離が予め定められた値以下の場合に実行される。上記距離が一定値以下の場合にのみマッチング処理を行うことにより、マッチング処理の回数を少なくして処理を効率良く行うことができる。

## 【0112】

マッチング処理において、最初に、当該 2 本の直線の端点と端点との間を結ぶパターンが生成される（ステップ S53-1）。即ち、マッチングに用いるパターンは、予め用意されるのではなく、その都度生成される。従って、多数のパターンを格納する必要がなく、そのための大容量のメモリも必要ない。パターンの生成には  $n$  次関数が用いられる。例えば、図 16（A）に示す状態は、横線と横線（又は縦線と縦線）の端点が近い位置に存在する例である。この場合、2 本の横線の端点  $a$ 、 $b$  に対するパターン  $a'$   $b'$  が生成される。生成されるパターン  $a'$   $b'$  は半円とされる。図 16（B）に示す状態は、縦線と横線の端点が近い位置に存在する例である。この場合、縦線と横線の端点  $A$ 、 $B$  に対するパターン  $A'$   $B'$  が生成される。生成されるパターン  $A'$   $B'$  は  $90^\circ$  の円弧とされる。生成されたパターン  $a'$   $b'$  又は  $A'$   $B'$  と画像パターンとが一致すれば（又は相似すれば）、当該角部は丸角部であるとされる。

## 【0113】

次に、生成されたパターンと入力された画像データ 12 上の画像パターン（入力パターン）とが比較される（ステップ S53-2）。この比較の結果が一致すれば、当該角部は当該生成されたパターンと同一の形状を有する丸角部であると

される。これにより、掠れた実線や間隔が広い点線についても正確に丸角部を決定することができる。

#### 【0114】

なお、当該画像パターンに近似させたパターン（近似パターン）を生成して、これと入力パターンとを比較しても良い。近似パターンは当該画像パターンを矩形近似することにより得る。近似パターンを用いることにより、入力パターンが点と点との間隔の長い点線や一部が掠れた丸角であっても、正確に丸角部を決定することができる。

#### 【0115】

以上のように、マッチング処理は、画素密度処理及び対象性処理によっては丸角部を決定することができない場合にのみ実行される。これにより、帳票の角部の形状（又はその規則性）に基づいて丸角部を決定する処理モードによっては、丸角部か否かを決定することができない場合（例えば、点線の場合）でも、丸角部を決定することができる。また、パターン生成処理の回数を少なくすることができるので、処理を効率良く行うことができる。

#### 【0116】

以上の処理の結果、図17に示すように、画像データ12から帳票の角部における丸角部が抽出される。これが丸角部データ16である。

#### 【0117】

図17は文字認識領域決定装置6の実行する文字認識領域決定処理について示す。この処理は、図3のステップS6であり、周知の処理である。

#### 【0118】

図17は、ある角部が丸角部決定処理において丸角部であると決定された場合における、当該丸角部とセルC1との関係を示す。図17に示すように、当該決定された丸角部はセルC1の内側に含まれた状態となる。なお、「a b c」は帳票に記入された文字である。そこで、文字認識領域決定装置6は、まず、セルC1の直ぐ内側の点線で示す矩形の領域を仮の文字認識領域C2とする。その後、文字認識領域決定装置6は、仮の文字認識領域C2から当該丸角部の存在する領域（斜線を施した三角形の領域）C3を除くことにより、真の文字認識領域を決

定する。これにより、当該丸角部が文字と認識されることを防止することができ、また、文字「a b c」を正しく認識することができる。

## 【0119】

以上、本発明をその実施の形態に従って説明したが、本発明はその主旨の範囲内において種々の変形が可能である。

## 【0120】

例えば、表画像処理装置100は上述した構成を全て備える必要はなく、適宜簡易な構成としても良い。特に、直線抽出装置2、丸角候補領域検出装置3、セル抽出装置4、丸角部決定装置5、文字認識領域決定装置6はソフトウェアによって実現されるため、CPUの負担が大きく、メモリも大きくなる。そこで、各装置の構成を簡易なものとしても良い。この場合、丸角部の抽出精度は多少劣るが、処理速度は高速化することができる。

## 【0121】

一例として、丸角候補領域検出処理における候補検出処理として重なり処理（ステップS32-1）のみを実行し、丸角部決定処理として画素密度変化処理（ステップS51）のみを実行するようにする。この場合、重なり処理のみが実行されるので、予め定められた距離に位置する縦方向の直線と横方向の直線との組（図8に示す例）のみが、丸角部の処理対象とされる。

## 【0122】

他の例として、上述の例において、丸角候補領域検出処理における候補検出処理として、重なり処理（ステップS32-1）に代えて、近接処理（ステップS32-2）のみ、探索処理（ステップS32-3）のみ、又は、これらの3つの処理から選択した2つの処理のみを実行するようにしても良い。丸角部決定処理としては画素密度変化処理（ステップS51）のみを実行するようにする。近接処理の場合には図9に示す例のみが丸角部の処理対象とされ、探索処理の場合には図10に示す例のみが丸角部の処理対象とされる。

## 【0123】

逆に、上述の例において、丸角部決定処理として、画素密度変化処理（ステップS51）のみに代えて、画素密度変化処理及び対象性処理（ステップS52）

、又は、画素密度変化処理及びマッチング処理（ステップS53）を実行するようにしても良い。候補検出処理としては重なり処理（ステップS32-1）のみを実行するようにする。

#### 【0124】

以上の説明から判るように、候補検出処理及び丸角部決定処理は、各々、上述した複数の処理から適宜選択して実行することができる。従って、これらの処理は、上述した例の他にも種々組み合わせることができる。

#### 【0125】

更に、他の例として、丸角候補領域検出装置3及び丸角部決定装置5のみを実施の形態において説明した構成として、他の各装置は上述した以外の周知の手段で実現しても良い。この場合でも、丸角候補領域検出装置3及び丸角部決定装置5の構成は上述した簡易な構成としても良い。

#### 【0126】

また、例えば、図2に示す処理フローにおいて、処理の順序を種々変更することが可能である。即ち、直線抽出処理（ステップS2）の後に、丸角候補領域検出処理（ステップS3）を行うことなく、先にセル抽出処理（ステップS4）を行ないセルの形状を決定し、その後に抽出したセルの角部について丸角候補領域検出処理（ステップS3）及び丸角部決定処理（ステップS5）をこの順に行うようにしても良い。逆に、直線抽出処理（ステップS2）及び丸角候補領域検出処理（ステップS3）の後に、セル抽出処理（ステップS4）を行うことなく、先に丸角部決定処理（ステップS5）を行って丸角部の形状を決定し、その後にセル抽出処理（ステップS4）を行うようにしても良い。更に、このような処理順序を変更することに加えて、上述したような丸角候補領域検出装置3及び丸角部決定装置5の簡易な構成を併せて採用することができる。

#### 【0127】

また、例えば、ある特定の処理の結果に基づいて、図2に示す処理の順序を種々変更することが可能である。即ち、ステップS31において抽出された全ての斜め成分Y及び斜め成分Tの大きさ（斜め線の長さ）が予め定められた値より小さい場合、候補検出処理（ステップS32）及び直線情報作成処理（ステップS

33) を省略して、セル抽出処理 (ステップ S4)、ステップ S31 を除く丸角候補領域検出処理 (ステップ S3)、丸角部決定処理 (ステップ S5) の順に実行しても良い。更に、この場合、セル抽出後の候補検出処理 (ステップ S32) を省略して実質的に丸角候補領域検出処理 (ステップ S3) を省略することができる。このようにしても、丸角部の面積が小さいので、直接丸角部決定処理を行う際の処理の負担が大きくならずに済む。

#### 【0128】

##### 【発明の効果】

以上説明したように、本発明によれば、表画像処理装置において、抽出した直線の端点を起点とする斜め成分を抽出し、抽出した斜め成分を用いた所定の処理により検出した丸角候補領域を含むセルの角部について所定の処理により丸角部を決定することにより、丸角候補領域を正確に抽出して丸角部を正確に決定することができるので、罫線に文字が接触する部分や画像の掠れた部分を丸角であると認識することを防止でき、罫線の掠れ、文字の丸い部分、罫線と文字との接触等があっても丸角を正確に処理することができ、この結果、罫線構造の認識及びセル (文字認識領域) の抽出を安定して行うことができる。

#### 【0129】

また、本発明によれば、以上に述べた処理を当該表画像処理装置であるコンピュータに実現させるプログラムを格納したプログラム記憶媒体を提供することができるので、丸角部を含む罫線構造の認識及びセル (文字認識領域) の抽出を安定して行う表画像処理装置を容易に提供できる。

##### 【図面の簡単な説明】

#### 【図1】

本発明の原理構成図である。

#### 【図2】

表画像処理装置構成図である。

#### 【図3】

表画像処理フローである。

#### 【図4】

直線抽出処理フローである。

【図 5】

線分抽出処理の結果を示す図である。

【図 6】

直線抽出処理の結果を示す図である。

【図 7】

丸角候補領域抽出処理フローである。

【図 8】

丸角部分を拡大した例を示す図である。

【図 9】

近接処理の例を示す図である。

【図 10】

探索処理の例を示す図である。

【図 11】

直線情報の作成処理を示す図である。

【図 12】

セル抽出処理フローである。

【図 13】

丸角部決定処理フローである。

【図 14】

画素密度変化処理を示す図である。

【図 15】

対象性処理を示す図である。

【図 16】

マッチング処理を示す図である。

【図 17】

文字認識領域決定処理を示す図である。

【図 18】

丸角表構造の一例を示す図である。

【符号の説明】

- 1 画像入力装置
- 2 直線抽出装置
- 3 丸角候補領域検出装置
- 4 セル抽出装置
- 5 丸角部決定装置
- 6 文字認識領域決定装置
- 11 入力画像
- 12 画像データ
- 13 直線データ
- 14 丸角候補データ
- 15 セルデータ
- 16 丸角部データ
- 17 文字認識領域データ

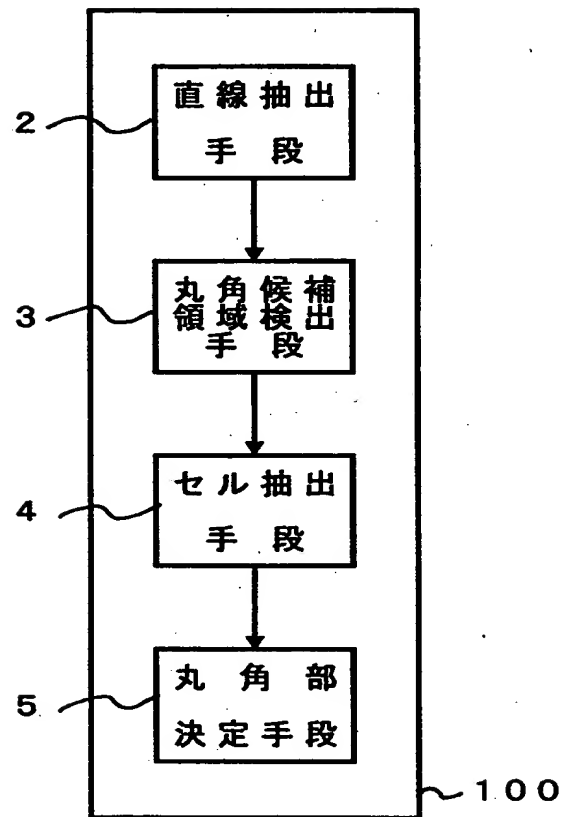


【書類名】

図面

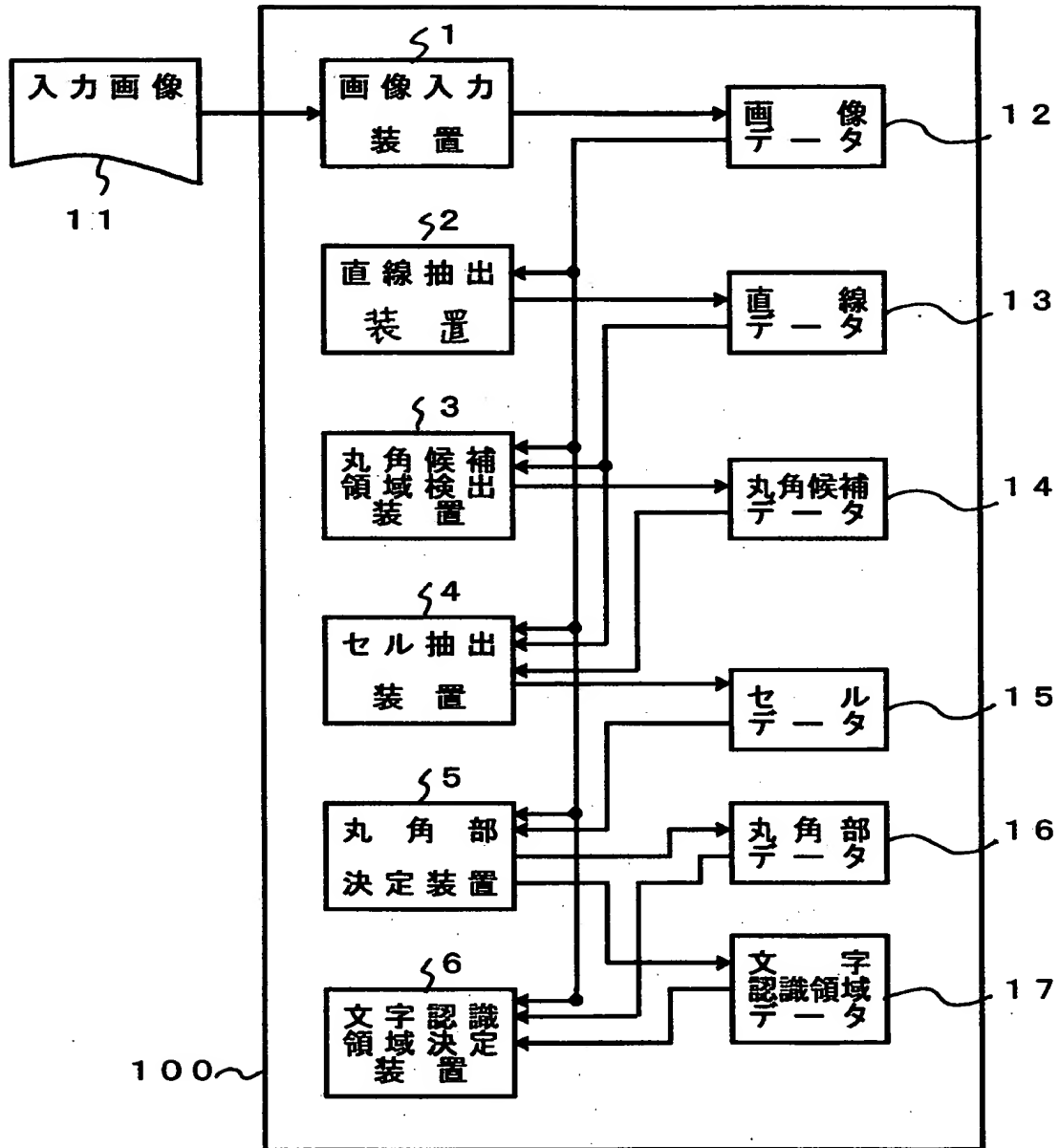
【図1】

発明の原理構成図



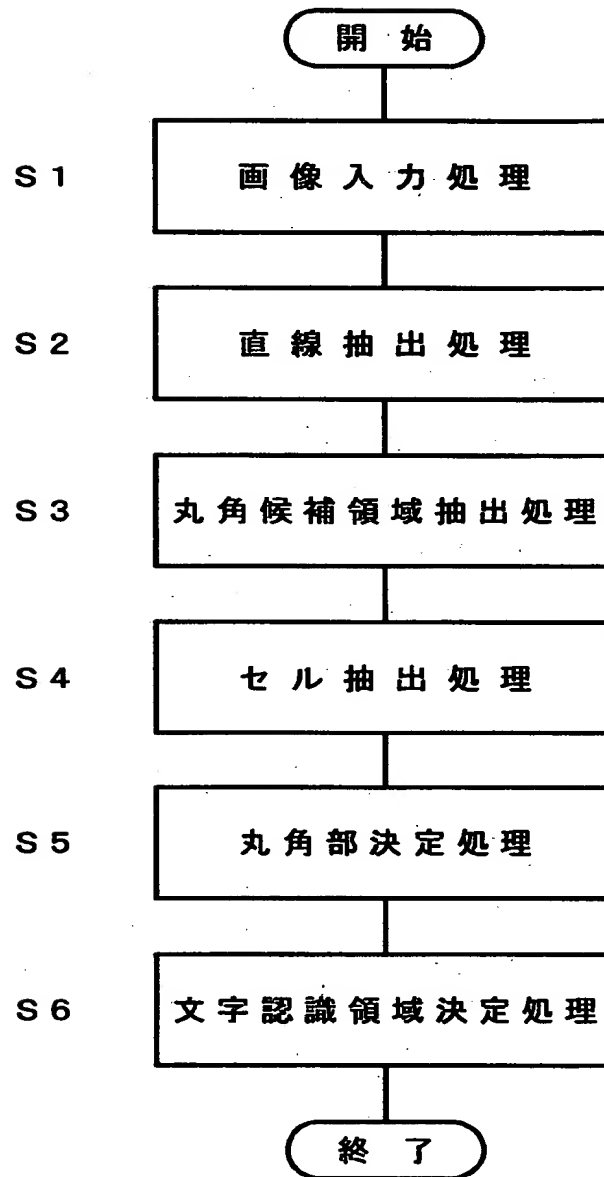
【図 2】

表画像処理装置構成図



【図 3】

表 面 像 処 理 フ ロ ー



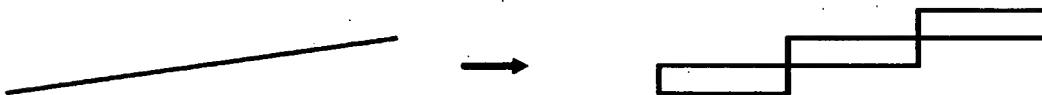
【図4】

直線抽出処理フロー



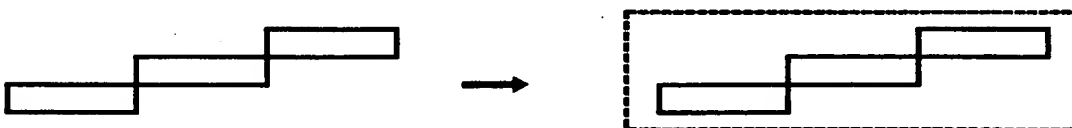
【図5】

線分抽出処理の結果を示す図



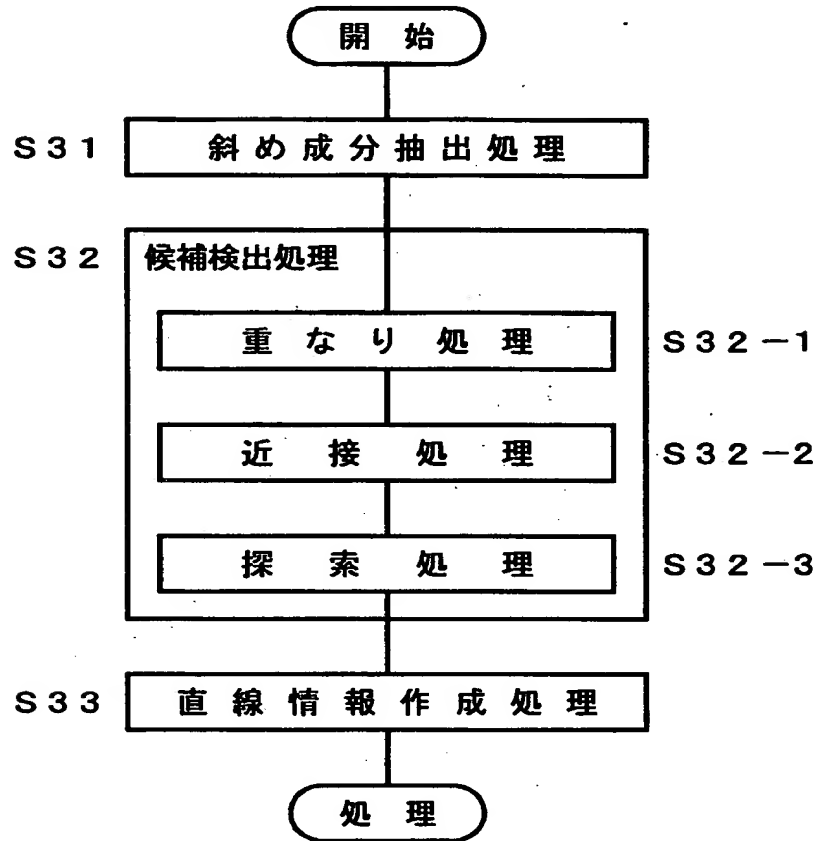
【図6】

直線抽出処理の結果を示す図



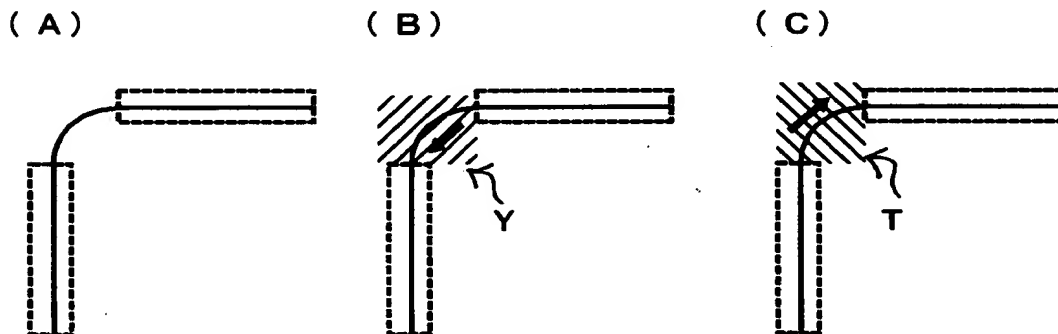
【図7】

丸角候補領域抽出処理フロー



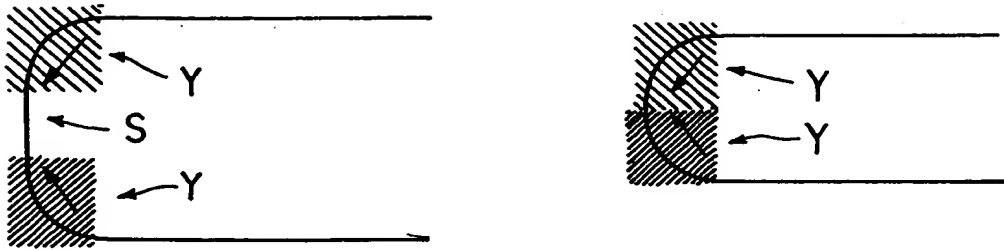
【図8】

丸角部分を拡大した例を示す図



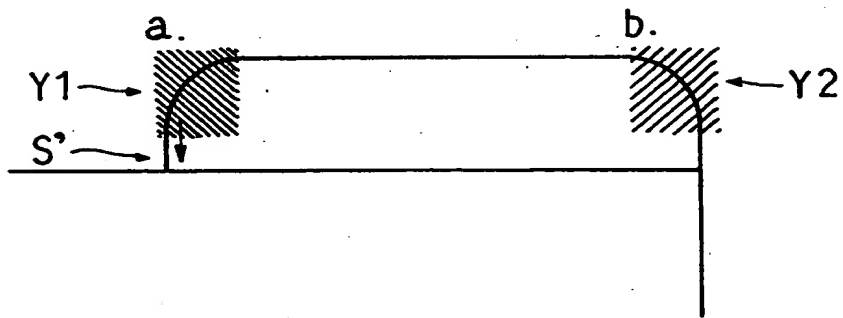
【図 9】

近接処理の例を示す図



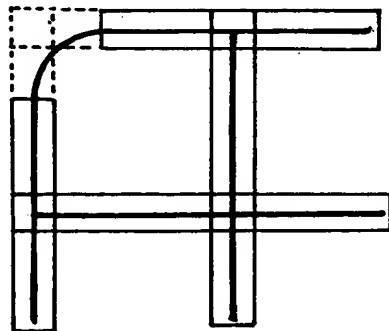
【図 10】

探索処理の例を示す図



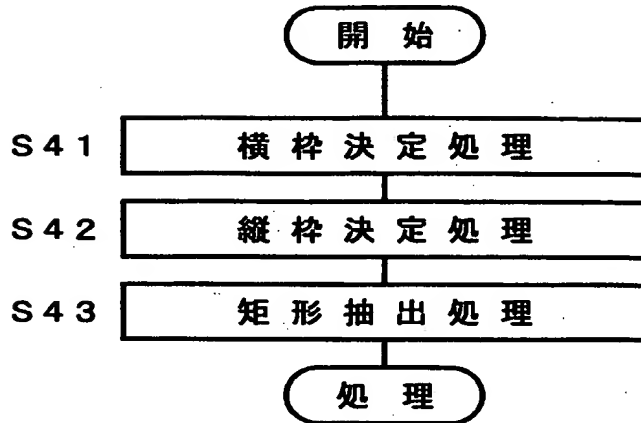
【図 11】

直線情報の作成処理を示す図



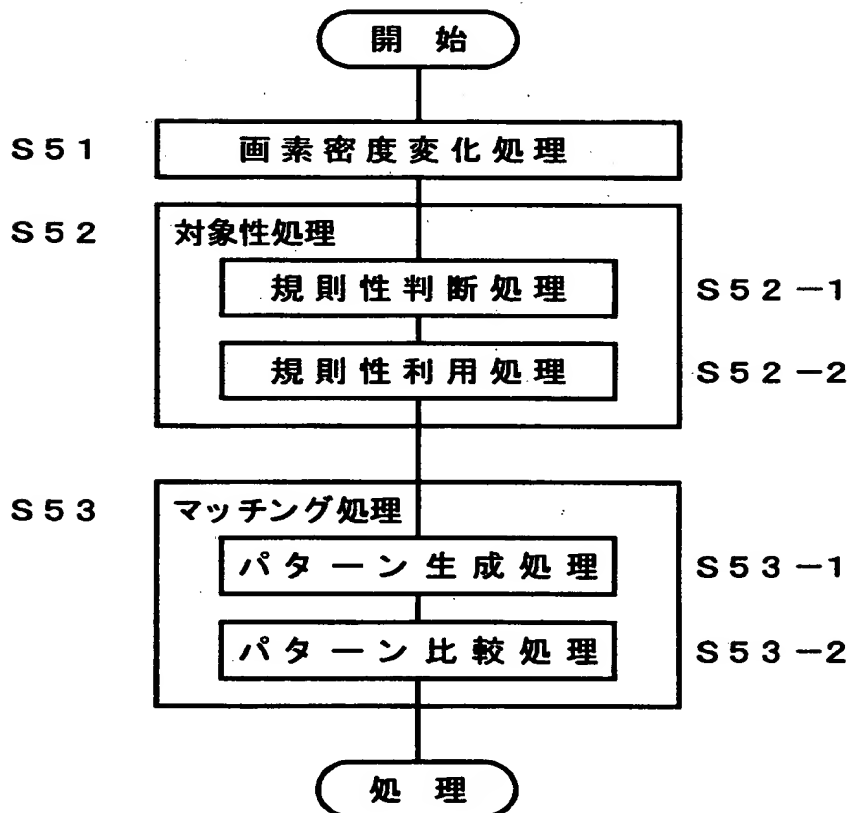
【図 12】

## セル抽出処理フロー



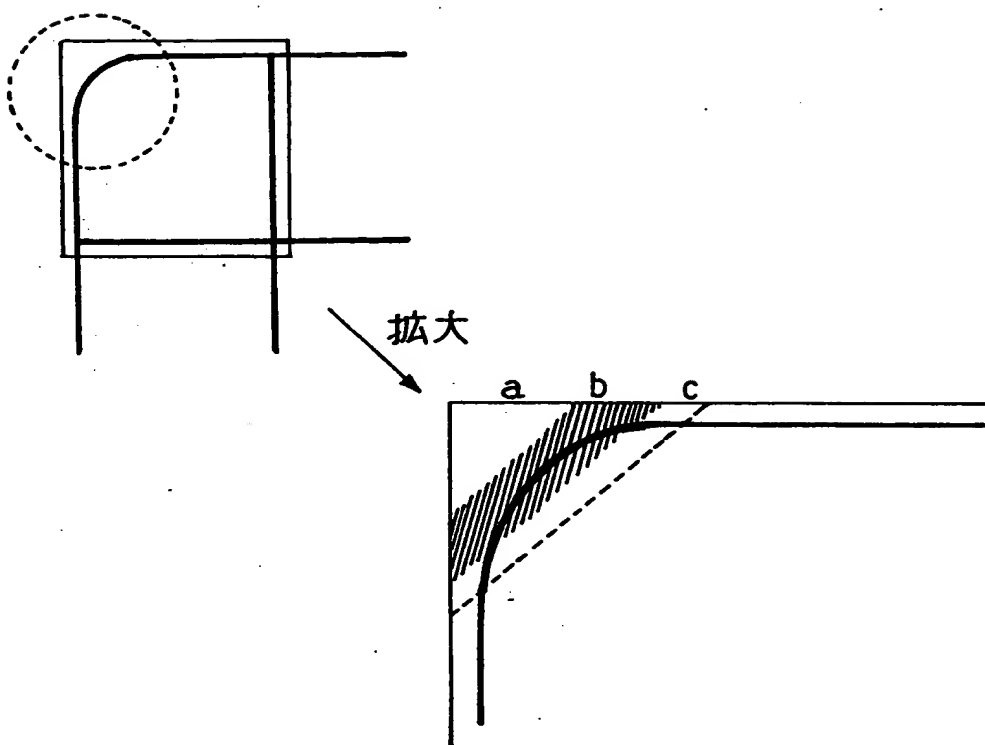
【図 13】

## 丸角部決定処理フロー



【図 14】

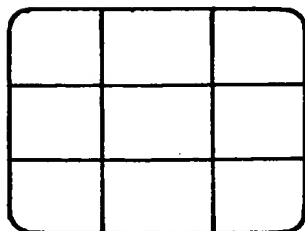
画素密度変化処理を示す図



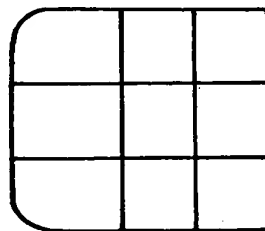
【図 15】

対象性処理を示す図

(A)



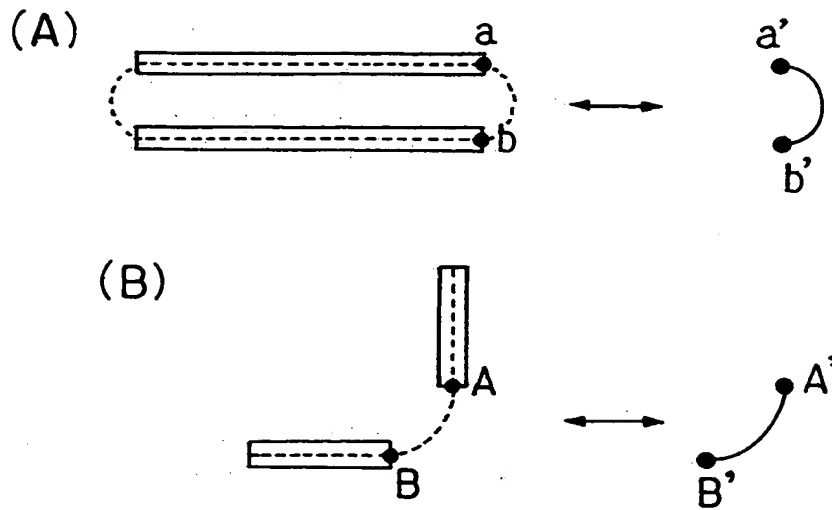
(B)





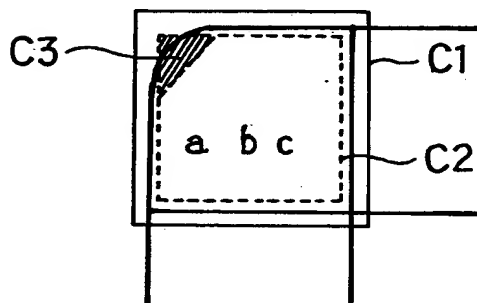
【図 16】

マッチング処理を示す図



【図 17】

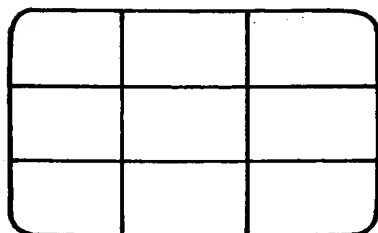
文字認識領域決定処理を示す図



【図 18】

丸角表構造の一例を示す図

(A)



(B)



【書類名】 要約書

【要約】

【課題】 本発明は、表面像処理装置及びそのプログラム記憶媒体に関し、丸角を有する表面像を正確に処理することを目的とする。

【解決手段】 表面像処理装置は、入力された画像から縦及び横方向の直線を抽出する直線抽出手段 2 と、直線抽出手段 2 の検出した直線を起点とする斜め成分を抽出し、抽出した斜め成分を用いて丸角候補領域を検出する丸角候補領域検出手段 3 と、丸角候補領域検出手段 3 の検出した丸角候補領域を含むセルを抽出するセル抽出手段 4 と、セル抽出手段 4 の抽出したセルの角部について丸角部を決定する処理を行う丸角部決定手段 5 とを備える。

【選択図】 図 2

【書類名】 職権訂正データ  
【訂正書類】 特許願

<認定情報・付加情報>

【特許出願人】

【識別番号】 000005223

【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号

【氏名又は名称】 富士通株式会社

【代理人】 申請人

【識別番号】 100087848

【住所又は居所】 東京都荒川区西日暮里5丁目11番8号三共セントラルプラザビル5階

【氏名又は名称】 小笠原 吉義

【選任した代理人】

【識別番号】 100074848

【住所又は居所】 東京都荒川区西日暮里5丁目11番8号三共セントラルプラザビル5階

【氏名又は名称】 森田 寛

【選任した代理人】

【識別番号】 100087147

【住所又は居所】 東京都荒川区西日暮里5丁目11番8号三共セントラルプラザビル5階

【氏名又は名称】 長谷川 文廣

出 願 人 履 歴 情 報

識別番号

[000005223]

1. 変更年月日

1996年 3月26日

[変更理由]

住所変更

住 所

神奈川県川崎市中原区上小田中4丁目1番1号

氏 名

富士通株式会社